DOCUMENT RESUME

ED 137 114 SE 022 340

AUTHOR	Salmon-Cox, Leslie; Holzner, Burkart
TITLE	Formative Evaluation as a Social Process: A Case Study.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	LRDC-1976-19
PUB DATE	76
NOTE	30p.; For related documents, see SE 022 339-341; Not available in hard copy due to small print throughout entire document
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	*Curriculum Development; Educational Research; Elementary Education; *Elementary School Science; *Formative Evaluation; Individualized Instruction; Instruction; Science Curriculum; Science Education
IDENTIFIERS	Learning Research and Development Center; Research Reports

ABSTRACT

This paper reports the results of a case study of the process of formative evaluation in the development of one unit of Individualized Science (IS). It is an analytic description of the curriculum project's experiences in formative evaluation rather than an evaluative critique. Following an introductory explanation, the IS program and the IS program directors' conception of formative evaluation are discussed. The concluding sections analyze the activities observed (both in the specific case and in general terms), the manifest and latent functions of formative evaluation, and some structural implications of these functions. (Author)

LEARNING RESEARCH AND DEVELOPMENT CENTER

1976/19

FORMATIVE EVALUATION AS A SOCIAL PROCESS: A CASE STUDY

LESLIE SALMON-COX AND BURKART HOLZNER

University of Pittsburgh

FORMATIVE EVALUATION AS A SOCIAL PROCESS: A CASE STUDY

Leslie Salmon-Cox and Burkart Holzner

Learning Research and Development Center

University of Pittsburgh

1976

The research reported herein was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred.

Abstract

This paper reports the results of a case study of the process of formative evaluation in the development of one unit of Individualized Science (IS). It is an analytic description of the curriculum project's experiences in formative evaluation rather than an evaluative critique. Following an introductory explanatory section, the IS program and the IS program directors' conception of formative evaluation are discussed. The concluding sections analyze the activities observed--both in the specific case and in general terms, the manifest and latent functions of formative evaluation, and some structural implications of these functions. The intended audience are those engaged in curriculum development, in formative evaluation, and other students of knowledge production and utilization.

FORMATIVE EVALUATION AS A SOCIAL PROCESS: A CASE STUDY

Leslie Salmon-Cox and Burkart Holzner

Learning Research and Development Center
University of Pittsburgh

The research reported herein comes from a case study of formative evaluation, a domain of activity characteristic of educational design and development. Formative evaluation is a mechanism to assess the adequacy of instructional materials and procedures as they are taking form. It is a technical tool for producing better curricula. In this study, however, we were not concerned with the substantive outcomes of the application of the process of formative evaluation. We were not concerned with whether or not the materials being evaluated were judged "good" or "bad" or something in between. Our focus was on the social processes of formative evaluation as they evolved in a structurally defined setting.

The authors of this report are sociologists whose theoretical frameworks are rooted in the sociology of knowledge. It is the intention of such sociology to understand both the mechanics of knowledge production and the meanings attached to those mechanics in the minds of the actors involved. In pursuing these objectives, our methods are similar to those of the anthropologist conducting field work.

There were three distinct purposes guiding this research study. First was the need for a fine-grained description of the formative evaluation process. What actually occurred during that time when the curriculum designers were engaged in development and prototype testing and simultaneously conducting evaluation they labeled "formative?" Second, given the activity description we constructed, what kind of comparison could be drawn between this and the developers' description of their own activity? Our third

purpose was the analysis of what had been observed to articulate the basic components of the processes involved and their interrelationship. This analysis dealt with three aspects: (a) problem-detection procedures, (b) decision-making procedures, and (c) the manifest and latent functions of the evaluation process.

Before discussing the details of the study, it is useful to describe briefly the organization in which the research took place and the role the authors play in the organization. The research was undertaken at the Learning Research and Development Center (LRDC) at the University of Pittsburgh. LRDC is an institute designed to bring behavioral science to bear on education for the purpose of improving practice in the schools. In particular, LRDC is concerned with designing environments for learning that are adaptive to the individual characteristics of students. As part of this work, the Center has developed a number of new curriculum programs for the preschool and elementary school.

As sociologists, the authors' major role within the Learning Research and Development Center is to study the organization as a whole and to report to an external Board of Visitors who monitor and guide the Center in its activities. In the conduct of studying events and trends in the Center, Salmon-Cox was attending a series of meetings being held to plan the multifaceted evaluation of Individualized Science (IS), an elementary school science program under development at LRDC. This innovative evaluation strategy, which encompassed studies done from several perspectives and resulted in an array of various kinds of information, is described by Leinhardt (1975), who planned and coordinated the effort. In November 1974, the IS project directors proposed that, as part of the larger evaluation effort, the authors conduct an intensive study of the process of formative evaluation as it is carried out by the IS project. It was proposed that we study the prototype test and formative evaluation of the "Quetelet" unit (a unit on statistical concepts) of the IS curriculum which was planned for Spring 1975.

Our research was conducted in the Spring and early Summer of 1975. As sociologists, we employed the methodology of observation, interview, document analysis, and attendance at all possible meetings, as well as conducted some in-classroom observation to collect data for the study. Though the formative evaluation itself had not been completed, sufficient data were collected to provide material for the present analysis.

In order to provide a context for our analysis of the processes of formative evaluation, we begin this paper with a description of the Individualized Science program and the procedures of formative evaluation as outlined by the IS project directors (Champagne & Klopfer, 1974). In addition to defining our domain of inquiry, their view of formative evaluation provides a mosaic to which later descriptions of actual activity can be compared.

In the next sections of the paper we present the data base for this research--the development of formative evaluation processes by the IS project. We analyze the activities undertaken, focusing on the social organization that performed the evaluative task, and discuss the manifest and latent functions of the particular formative evaluation under study. In the concluding section, we reflect upon the total process observed, on the functions of formative evaluation in general, and on some structural implications of those functions. We turn now to an overview of the Individualized Science program.

The Program Under Evaluation: Individualized Science

Audrey Champagne and Leopold Klopfer are co-directors of the Individualized Science project and co-developers of IS. In their paper entitled "Individualized Learning in Individualized Science" (1975), they describe the program as follows:

> Individualized Science (IS) [is] an elementary-school science program . . . designed to serve children in present school grades K through 8.

> There are seven levels of the IS program, each of which consists of approximately one year's work in science. The science content of IS is drawn from both the biological and the physical sciences. One unique aspect of the program is its emphasis on the cultural aspects of science and the interactions of science and society. IS is a comprehensively individualized elementary-school science program. Not only does IS provide mechanisms for a child to help plan his or her science activities, to manage his or her own instructional materials, and to take part in the assessment of his or her learning, but it also provides opportunities for the child to make selections from among alternative learning resources and from among alternative units of study. (p. 1)

The authors specify five goals for the IS program. These goals are:

1. Scientific literacy goal: The student acquires a foundation for scientific literacy.

2. Student self-direction goal: The student views the learning process as primarily self-directed and self-initiated.

3. Student co-evaluation goal: The student plays a major role in evaluating the quality, extent and rapidity of his learning.

4. Affective goal: The student displays informed attitudes toward his study of science, scientific inquiry, and the scientific enterprise.

5. Inquiry goal: The student becomes skillful in using the processes of scientific inquiry and is able to carry out inquiries. (pp. 7-8)

In discussing these goals, they point out that three of the program's goals are related to learning science content and two are related to the student's management and evaluation of his or her own science learning. "Each of the five goals is subdivided into level competencies and the science subject matter learnings are specified in unit behavioral objectives. The analysis of goals into competencies and behavioral objectives has guided the development of the learning resources of IS" (p. 8). The co-developers go on to describe 19 different kinds of learning resources in

IS which differ in "function, components, instructional strategy, primary sense modality, and the social setting in which they are used."

In the Quetelet unit on statistical concepts, the students collect and record data and perform some statistical analyses. During the prototype testing phase of this unit, a number of learning resources were being evaluated: Readings in Science (RIS), illustrated booklets containing several readings, optional read-along tapes, with comprehension and extension questions accompanying the readings; Invitations to Explore (ITEs), illustrated booklets with printed texts, optional read-along tapes, manipulative materials, and a response sheet booklet; Student Activities (SA), printed cards which describe activities, practice sheets, and manipulative materials. These resources in the Quetelet unit are referred to in this paper in abbreviated form.

Formative Evaluation in Curriculum Development: The Rationale

An article entitled "Formative Evaluation in Science Curriculum Development" written by Champagne and Klopfer (1974) sets forth their philosophy and outlines procedures for formative evaluation. They stress that evaluation is an integral part of the development process. Their philosophy stems from the distinction made by Cronbach (1963) and Scriven (1967) between "formative" and "summative" evaluation. Formative evaluation is carried out while the program is being developed and the findings of the evaluation serve as a basis for revision of the program. Summative evaluation takes place after the program "goes out to the public" and changes based on the findings of this evaluation cannot be made easily.

Champagne and Klopfer state that formative evaluation must determine the effectiveness of the program--the outcomes which are explicitly planned. But questions of effectiveness, the focus of most evaluation studies, "are far from being the only ones that must be considered." They

believe formative evaluation must also determine the effects of the program--the outcomes which are not among the specified aims.

The authors propose four stages of formative evaluation which span a continuum from the initial conceptual phase through the final implementation and teacher preparation phase. During Stage A, IS staff members focus on conceptualization and planning and it is during this stage that the instructional materials are designed and reviewed before being tried out in a classroom. During Stage B, the prototype testing stage, the materials are tried out in one of LRDC's associated developmental schools. The Stage B evaluation is conducted by school personnel and members of the IS development group and revisions are made based on the information they obtain. In Stage C, the materials are implemented and evaluated in a network of field test schools. Stage D evaluation centers on the effectiveness of the teacher preparation program, the ease with which the program is implemented, and marketability. Data from Stages C and D are not collected by IS or LRDC staff but by Research for Better Schools (RBS), an educational laboratory located in Philadelphia which has been responsible for field testing and dissemination of several programs developed at LRDC. It should be noted that Stages C and D are different in character from A and B which involve only LRDC staff and a classroom in an associated school.

The authors have developed a series of questions to be addressed at the varying stages. The questions are organized into six categories, each of which focuses on a different aspect of the formative evaluation: (a) conceptualization and planning, (b) student instructional materials, (c) student behaviors with respect to the science content, (d) student behaviors in the classroom environment, (e) the teacher, and (f) marketability. The authors admit it may be impossible to address the multitude of questions they have posed (as many as 40 in one category) in any single formative evaluation. But they suggest that an adequate sampling of each of the questions in each category is essential. They point to three criteria for question selection: One should probe the most innovative aspects of the program, the aspects

of the program which its developers think most significant and interesting, and those areas which provide relevant data, the collection of which is not too difficult.

Perusal of the multitude of questions shows that they do indeed span a wide range of program effects. In addition to the more usual questions regarding content selection and instructional strategies, there is also a great deal of emphasis on inquiry skills, self-management and self-direction, and the overall culture of the classroom as it is created by and contributes to the program's functioning. The students' and teachers' feelings about the materials are considered along with the more pragmatic concerns of packaging, efficient storage, costs, and such.

In their article, Champagne and Klopfer appear to posit role differentiation between designer and evaluator; that is, they do not seem to expect that these two functions will be conducted by the same person or people. Potential problems resulting from this role differentiation are anticipated--problems of responsibility and relative authority. They discuss, from their point of view as designers, the kind of relationship that they feel must exist between an evaluator and a curriculum designer. They state that an evaluator should present data "in a thoughtful way, a way that is honest and yet does not threaten or antagonize the developer." The authors consider the findings from formative evaluation confidential. The formative evaluation "should remain privileged information with the curriculum development organization, where it can be meaningfully interpreted." They note that the "authority to make modifications in the program and to decide on the nature of the modifications resides with the developers" (1974, p. 202).

It is interesting at this point to note that the developers of IS partially solved the role differentiation problem by designing and conducting their own formative evaluation at Stages A and B (as they have defined these). At later stages, they keep a close watch over the evaluative activities of others.

The Development/Evaluation Process

The development of the Quetelet unit, from conceptualization until prototype testing, covered a period of about two years. Our study of the IS project began late in that time, but our research allowed us to reconstruct events and the involvement of salient actors from the unit's inception. Such a reconstruction was essential for our analysis, placing what we could observe into the larger context of what had occurred from the start.

The Quetelet unit is a mainstream--i.e., mandatory, basic core--unit of Level G of IS. This implies that it will be used mainly by seventh- and eighth-grade students; however, since IS is an individualized program, students at other grade levels can work in the unit. The bulk of the curriculum materials for the unit were written in 1973. In 1974, with editing and rewriting, the instructional materials were completed. In April 1975, when the prototype version was produced by the publisher, implementation of the unit began in an eighth-grade classroom at St. Mary's school, a Center-associated school located in Pittsburgh.

Expectations Concerning Evaluation of Quetelet

The general consensus among those who saw Quetelet before its use in the school was that this was one of the best prepared, most polished units ever to be entering prototype testing in IS. There was, however, one anticipated source of problems with the content. This was the question of the level of mathematical sophistication needed to work in Quetelet. Designed to teach basic statistical concepts to seventh and eighth graders, the unit appeared to require an ability and ease with numerical manipulation which some doubted the students would possess. Klopfer dismissed this possibility, saying he thought there was a good match between the students' mathematical level and what was necessary for Quetelet.

What was found on this question is an interesting example of the constraints placed by the "real world" on curriculum design, testing, and evaluation. The anticipated problem during the prototype test at St. Mary's could not be investigated because the bulk of St. Mary's eighth graders who were working on Quetelet had a strong background in the major concepts of the unit. Informal discussion among the school and IS project staff involved seemed to indicate that everyone suspected that this kind of student background was out of the ordinary. Eighth-grade mathematics classes do not normally cover statistical concepts. However, of the 18 students to work on Quetelet in the Spring of 1975, 13 were in a "high math track," and found much of the substantive content of Quetelet to be a repeat of work done in their mathematics class in the same academic year. The remaining five students were in an average mathematics class, and they, too, seemed familiar with many of the concepts.

A course of action was outlined to remedy this problem. They planned to collect further data in the fall on a new group of "low mathematics" eighth graders and on seventh graders, with the hope, or expectation, that these groups would not be prefamiliar with the content. Also, the people at RBS responsible for the field testing of Quetelet would be alerted to the possible occurrence of difficulty with the mathematical level, as well as to the apparent inability of the Pittsburgh prototype test to investigate this question.

Another more general anticipation was that because Quetelet was in "prototype" form, it would pose certain problems. This worry was expressed by the assistant teacher, Karen Evans, who had not been involved in prototype testing before. Since the unit had never been in the classroom before, she feared it might be in very rough form. In particular, she was concerned that the necessary supplies would not arrive on time, the printed materials would not be clear, and something might go wrong with the manipulables. Sister Martha, the teacher at St. Mary's had been involved in prototype testing of another unit of the IS program and had experienced a

number of problems. It would not, therefore, have been surprising if she harbored some fears about working through Quetelet also. If she did, however, she did not express them to the observer.

By mid-May, after about a month of implementation, Evans reported to the observer that the problems expected were not appearing. Not only were there none of the anticipated problems with the unit, but no unanticipated trouble spots had arisen either. The prototype materials of Quetelet were in order; in fact, she found that she had more problems with "already done" units of IS. Because of the students' background in math, her expectation that she would be required to provide a great deal of explanation was not being realized. Things were going so smoothly, she said, that it was "almost boring." The children were highly motivated. They were eager to work in the unit to help out the people from the University, and they also seemed to enjoy the content.

Anticipations of the Formative Evaluation Session

In April, the IS staff began collecting observational data on students working in Quetelet. The paper by Champagne and Klopfer (1974) on formative evaluation was reviewed, and several groups of questions which would guide observation were selected. The questions were selected with reference to three criteria: (a) the answers to the questions were obtainable through observation; (b) the questions were relevant to observation of a single unit, rather than the whole program; and (c) they were appropriate to this phase of student use, namely prototype testing. Data collection proceeded through April, May, and part of June.

On July 7, 1976, a meeting was held at LRDC to discuss the Quetelet unit, bringing together the directors of IS, the IS staff responsible for collecting observational data during the evaluation, the classroom teacher, the assistant teacher, and a student from St. Mary's eighth grade who had completed the Quetelet unit. This meeting was a focal point of the formative evaluation of the unit.

People who had been involved in previous formative evaluations were asked how the long review meeting at LRDC was conducted and what usually occurred. There had not been many such in the past year or so. The only fairly recent one had been of the IS "Harvey" unit, where materials had undergone some significant changes.

In the Spring, Sister Martha's comments about the upcoming session on Quetelet were very general. She was vague about what details she might be questioned on, but did say that in the past she had critiqued the materials for consistency in instructions given to the students and clarity. She phrased her anticipated contributions primarily in affective terms, putting little emphasis on any role she might play in content evaluation. She said that the LRDC staff always seemed highly receptive and grateful for whatever comments she made. However, she did convey a picture of a session in which she and Karen Evans would play major roles, reporting on their classroom experience with Quetelet. Evans, having only one year's experience with IS, had no specific expectations concerning the July 7th meeting.

LRDC staff were also fairly imprecise in their expectations. They also gave the impression that Sister Martha and Karen Evans would have the major roles in the meeting. Therefore, what generally seemed to be expected by the participants was an informal meeting consisting essentially of questions asked by LRDC staff and answered by school personnel. Supplemental information would be provided by IS project staff observers when necessary.

Formative Evaluation Session

The July 7th meeting was held in the midafternoon and lasted for about two and one-half hours. Contrary to expectations, it was fairly formal, structured, and run by an LRDC staff member, Joan Fogarty. Fogarty was involved in some of the first writing for Quetelet and was in charge of the unit's formative evaluation. Inputs from the two teachers were minimal;

from IS staff assisting Fogarty in collecting observational data, there were none at all.

A number of factors set the tone for the meeting and probably made the discussion somewhat atypical. First, insufficient data had been collected before school closed, so this session was not really a "wrap-up," a final formative evaluation occasion, but rather an interim meeting. Second, it was clear to all concerned before the meeting was held that Quetelet was a fairly unproblematic unit. It was well written, had been working smoothly in the classroom, and seemed to require only minimal revisions. In addition, the presence of an eighth-grade student probably also affected the tone. Apparently, student participation had been talked of before, but this was the first time it actually occurred.

Joan Fogarty opened the meeting and spent the first hour discussing the preliminary nature of the evaluation and the questions that had guided observation. She said that the students had been observed and talked with in an informal, participatory way. She shared with all present the anticipated problems of the unit's mathematical level, and the fact that this could not be evaluated because of the familiarity most of the students had in the mathematical concepts of the unit. She also pointed out that a number of instructional materials--many of the RISs and SAs and some of the ITEs--had not been used by any students and some by only one or two. This fact, coupled with the small number of students and the homogeneity of the student sample, made the data incomplete. Since the mathematical level of the unit had been the major anticipated source of difficulty, and since data on it were inconclusive, it served as a background for the discussion which followed, being referred to in several contexts.

For the remainder of the meeting, conversation ranged over details of the unit. There was discussion of specific ITEs, and the meeting ended with some general talk of "where do we go next." The specific details discussed uncovered several points for revision, none of them implying radical change. Table 1 lists the specific problems raised at the meeting, indicating

Table 1

Outcomes of the LRDC Formative Evaluation of the Quetelet Unit of Individualized Science

	Suggested Problem	Source of Suggestion	Recommendation
1.	Nonuse of some manipulables.	Fogarty observations; Sister Martha and student concur.	Leave them in, as other students may find them useful.
2.	Lack of enough "wet" in-laboratory activities--activities which provide students with experience measuring. (The data they collect from measuring various items are used in learning the statistical concepts of the unit.)	Fogarty questioning of students.	Perhaps include more measuring activities.
3.	The "log book," a central activity of the unit in which students record and share data from measurements they have performed, had the following problems: (a) Fast-moving students required data from the log book before students responsible had collected it; (b) there were not enough items to measure and record; (c) some items specified for inclusion in the log book were too hard to measure.	Evans's observations; Fogarty and Klopfer concur.	Use "canned data," add more items to be measured, and perhaps use more easily measured items.
4.	Lack of enough specified student-teacher conferences.	Fogarty's observations; also student's comment during discussion.	Perhaps add more conferences and make optional; make explicit students' need to understand material before continuing.
5.	Difficult vocabulary.	Fogarty looked over materials beforehand, chose potential trouble-makers, confirmed many with students in conversation.	Change known trouble words, or write additional explanations.
6.	"Relevance" of unit to real world concerns.	Fogarty's observations and conversations with students.	None at this time. Review problem again when more students have completed the whole unit.
7.	Placement test not adequately keyed to instructional materials, i.e., test unduly easy.	Klopfer formulated problem after discussion.	Let RBS evaluate the placement test carefully; perhaps analyze how various students performed in the unit to determine more precisely what they need to know before beginning the unit.
8.	Student difficulty with the following ITEs: prediction, probability, and the normal curve.	Fogarty's observations; also general discussion.	Try out some alternative teaching strategies; then review again to decide whether or not to revise printed text.

how they were discovered, and the recommendations put forth by the evaluation group for their remedy. A brief review of three of the problems raised will give a feeling for how they emerged in the course of the meeting and how the respective solution related to the goals and philosophy of IS.

The first problem listed in Table 1--the nonuse of some manipulables--was first raised by Fogarty; Sister Martha and the student agreed that the problem had occurred. The reason given for their nonuse was that the students either completely understood or could mentally visualize the concepts being taught without props. Champagne suggested "dropping these manipulables." Discussion followed. Klopfer summed up the discussion and the result was a decision to leave them in the unit because other, less-prepared students might find them useful.

The decision to leave the manipulables in the program was an interesting one, the reason for which can be found in the underlying philosophy of the developers. The developers view IS as a multifaceted program that matches instruction to the various characteristics of individual learners (Champagne & Klopfer, 1975). Thus, IS has been designed so that each unit has a wide variety of learning resources (e.g., illustrated lesson booklets, audiotapes, exploration kits which contain manipulative materials and simple laboratory equipment, etc.) to try to assure maximum accommodation to individual differences. In the face of disconfirming evidence of the utility of some of the manipulatives--not a single student had used them--the developer's notion of their place in the instruction was unshaken.

The fourth problem--the lack of enough specified student-teacher conferences--was raised by the student at the meeting in response to questioning about what she would want to see done differently in the Quetelet unit. The problem was discussed at length as it had been a source of difficulty before with respect to earlier units. The publisher of the IS program had asked that more conferences be added to earlier units because as Champagne pointed out, the publisher felt that everything should be highly specified. Champagne said that in IS, "we want to give kids and teachers

suggestions." The number of student-teacher conferences is "the kind of decision a teacher makes." The resolution of this problem was twofold: Specify some additional optional student-teacher conferences, and strongly encourage students to seek help whenever they need it.

A final interesting point stemming from the discussion of problems had to do with the "relevance" of Quetelet, the sixth question in Table 1. The seventh and eighth graders' inability to perceive much relevance in statistical concepts is not totally surprising when we consider college student plaints on the same subject. However, the materials in Quetelet are specifically designed to make "real world" applications more obvious. The developers' decision was to return to this question when more students had worked their way through the entire unit.

In the closing minutes of the meeting, Sister Martha and Karen Evans both spoke of the aspects of Quetelet that they particularly liked. Sister Martha pointed out that Quetelet's activities gave students a chance to "find out what's behind what comes out of a science laboratory." Evans said that the emphasis on data recording and data collection helped to tie together much of the preceding instructional material of IS. There was general discussion of the great intellectual sophistication of contemporary seventh and eighth graders, making more salient the need for individualized programs of high quality such as IS. The meeting closed with consensus on the need to examine further the specific problems raised regarding particular instructional materials and teaching strategies and the student-teacher conference question.

Analysis of the Process of Formative Evaluation

A major concern expressed by Champagne and Klopfer (1974) in their paper on formative evaluation was the role distinction between designer and evaluator. In the formative evaluation of Quetelet, such a distinction did not exist for the IS project during the stage of prototype testing. Many of

the IS staff were involved in this evaluation. There was no member of the staff whose sole responsibility was the evaluative activity.

The model or structure of procedures which they present in their paper was applied by Klopfer and Champagne in the conduct of the actual formative evaluation of Quetelet, modified by the cognitive styles of the actors themselves. Although they presented groups of questions and stages of activity as a model, it is apparent that, in fact, a far more holistic approach was taken than is conveyed in the paper. By holistic we mean an approach that places greatest emphasis on total configurations, on how components fit with one another to make a coherent whole. This might be contrasted with a model of evaluation or of activity which assesses the precision with which any individual piece fulfills its particular objective and assumes that a collection of such individual "excellences" makes for an excellent whole.

This notion of a holistic approach is a consistent theme which permeates all activities connected with IS--curriculum design, internal project organization, project/other agency interaction, and evaluation of materials. There is more of an emphasis in IS materials on process than on specific content. Anyone who has looked at the IS materials cannot but be impressed with the quantity and complexity of science content in this elementary school program; "scientific literacy" is the first of the program's five goals. However, the other four goals emphasize style of interaction among students, teachers, materials, and the learning environment rather than mastery of science facts.

The IS program is considerably more than a sequence of specific learnings in science. It is a model for interaction and for a cultural environment which fosters self-initiative and emphasizes the aesthetic aspects of science. This emphasis on the total configuration of the program was evident in the evaluation process as well, where concern was more for congruency of the unit with the classroom and with the entirety of IS than for

specific details of content learning. This is not to say that specific content is overlooked. It is carefully scrutinized, but as a means to a more complex end.

Because there was so little "wrong" with Quetelet, it is difficult to give examples of this approach using that unit. However, a previous experience with formative evaluation on the IS Harvey unit (a biological unit on circulatory systems) is illustrative. As part of Harvey's resources, there was a series of Min-Exs--activities which required that the students build physiological models of certain biological systems. During formative evaluation, it was decided that these Min-Exs would have to be replaced. Klopfer reported that the activities were "dull," an assessment agreed to by the teachers and the LRDC staff. He said that on examination these Min-Exs were, in and of themselves, perfectly all right but that the children did not get excited over them. More importantly, they did not support the overall philosophy of IS (especially the emphasis on inquiry and self-direction) as well as they might have. New ones were developed.

Examination of the old and new Min-Exs reveals that though both are model-building activities, the new ones stress human rather than animal physiology. More importantly, the new ones are much less "directional." As students build the models, they are not simply told what to do, but are led through a series of "thought" questions about what they are doing and about the workings of the system they are modeling. In addition, once the models are built there are subsequent activities in which they are used, e.g., teaching another student a concept using the model. These subsequent activities were not part of the first set of Min-Exs.

Klopfer has said that the IS project's evaluative criteria in such cases are three: The materials must "appeal to the kids," match our goals, and meet an "indefinable kind of test," namely, relationship of any part to the total program.

This holistic approach is also visible in the internal project organization of IS. The project is led by two co-directors. Observation and staff

interviews reveal a consistent form of leadership--a division of tasks by time rather than task attribute; that is, both project leaders are involved in all phases of IS activity but one on some units, one on another. They do not duplicate each other. Decision making is perceived, and appears to an observer, to be totally mutual and by agreement between the two. Staff cohesion is a matter of some emphasis, with the project directors attempting to convey a particular style or model of activity.

Problem-Detection Processes

As we have seen in examples from both the Quetelet and Harvey units, the processes by which problems are detected are very similar to those described in the Champagne and Klopfer paper (1974). They are somewhat "stagelike." Initial problem detection is accomplished during conceptualization and early writing ("Stage A"). At that stage, detection appears to be based on some inner mechanism of the project co-directors, derived from their cumulative experience. As the students work with prototype materials ("Stage B"), they are observed, talked with, and their written work is analyzed. There is a marked absence of reliance on test scores to reveal much about the program's effects up to this point. Instead, the developers seem to rely upon gathering information from as many of the involved actors as possible.

The session at LRDC was itself a search for evidence that might disconfirm the initial successful set of experiences with the unit. There was a high degree of consensus in the meeting on Quetelet, regarding both what was "good" and what required some revision. The source of this consensus may be found in the commonality of values of those involved. Though perhaps varying in degree by role, all those at the session shared a value set which emphasizes the quality of a student's experience in a learning situation. This concern for people, for the human as well as the technological aspects of science, is also exemplified in the curriculum. The choice of a scientist's

name to title each unit is a significant one, and a contrast to the norm. IS's units are entitled Quetelet and Harvey, not Statistics and Circulation.

In Sister Martha, IS has a teacher whose values are congruent with those of the developers, so that consensus was achieved by agreement rather than by authority. However, it must be said that IS is functioning well (as defined by acceptance and maintenance of the program) in a large number of schools with a diversity of teachers. Even in the prototype stage, IS has been handled by teachers with several different value systems. As Champagne noted based on her observations in several prototype settings, IS appears to have enough flexibility to suit the needs of a variety of teachers. Consistent with the IS philosophy, if a teacher wished to strongly emphasize scientific literacy he/she would have the material for it in IS and could create that kind of environment.

Problem-Solving Processes

Problem solving in curriculum design is a creative process. It is a recasting, reformulating, or recreating activity which when completed results in modification of program content or teaching strategies. In the initial stage, problem solving is effected through editing. A great deal of this form of problem solving takes place before the prototype materials are completed.

Problem solving following prototype testing involves several strategies. Once a trouble spot has been found, it may be decided to: (a) make no alteration, as the rationale for inclusion-as-is still appears sound; (b) make some alterations, anything from minor editing to a major change (e.g., the total rewrites of the Min-Exs in Harvey); or (c) withhold decision pending more student experience with the material.

In the prototype testing of Quetelet, each of these strategies was used in relation to some particular problem. It was decided, as in the case of the nonuse of some manipulables, to retain some instructional elements

which could conceivably have been dropped. It was decided to make some changes in the management of Quetelet, such as specifying the timing of student-teacher conferences. In other places, vocabulary was changed or its context modified to enhance clarity. Finally, some questions (e.g., the one concerning students' appreciation of the relevance of the content) were left open pending more student work in Quetelet.

Manifest and Latent Functions of Formative Evaluation

This entire process of formative evaluation and most especially the focal part of it, the meeting at LRDC, serves both manifest and latent functions for those involved. At the manifest, or overt level, there is first the stated purpose. Formative evaluation helps in the redesign of curriculum materials and procedures before these are field tested. Many changes have been made in past units of IS based on formative evaluation results; even in the unproblemmatic Quetelet unit, some changes were recommended. Additionally, where there is a question, and the decision is made not to alter the material, this too serves a function. Should the problem reappear in field testing, and the evidence for change still not be considered persuasive by the co-developers, then the forethought and rationale for inclusion worked out in formative evaluation might well stand them in good stead in their dealings with others in outside agencies.

This, in fact, is a second manifest function of the process. That is, the IS project must deal with two outside agencies before materials are finally completed: RBS, the field test agency; and Imperial International Learning Corporation, the publisher, distributor, and marketer of the finished product. That a careful in-house review is conducted, involving both the classroom teacher and evidence of students' work in the unit, strengthens the hand of LRDC staff in working with these two outside organizations. Both RBS and Imperial have on their staffs people with some background in science education who propose modifications in IS. In negotiating

the form of the final product, the LRDC staff frequently need to marshall support for their position, and formative evaluation results are useful in this regard.

In addition to these overt functions, formative evaluation serves several latent functions. As a social process, it serves to validate and confirm the models for action which the IS project co-directors have laid out. By "models for action" we mean both the overall instructional model of IS as well as the model for behavior within the project and in the project's dealings with other actors and groups. This model conceives of activity as holistic. In the analysis, congruency of parts takes final primacy over excellence of pieces; all actors take several roles and become involved in several aspects of activity.

It is possible to conceive of an alternative, "atomistic" model of behavior. In a project so structured, one might expect to find a highly differentiated division of labor, with an accompanying set of highly specified roles arranged in a clear authority hierarchy. Interaction within the project and between project and other groups would tend to be fairly formal and specific. Similarly, curriculum design and evaluation would place primary importance on the quality of individual pieces developed rather than overall program configuration.

This is not the way the IS project functions. The project directors have a strong, shared model of appropriate action and it is their style of action which was modeled in the long evaluation session, with its many participants. There appears to be emphasis on the IS staff adopting this model. For example, in a hypothetical case in which a staff member wished to rely heavily on analysis of pre- and posttest scores, one could predict that this would be found unacceptable and that the rationale behind such an argument would be forcefully rejected.

The formative evaluation process inspires self-confidence among the actors involved in the several roles that they play. Regardless of the

amount of revision called for, the process itself tends to reinforce the positive aspects of development, i.e., "we-have-a-good-thing-going-how-can-we-together-make-it-better?"

For the classroom teacher, a special purpose is served by this process, namely, the provision of a peer group. To be a science teacher in an elementary school implies, in almost all cases, that the teacher is quite alone, at least in terms of having anyone with whom to share science education concerns. Especially in the case where the teacher demonstrates a willingness to be innovative--as she/he does by taking on a program under development--there is increased need for professional reinforcement. Meeting at LRDC and being asked for advice and guidance increases the teachers' professional self-concept.

Finally, fulfilling another latent function related to a holistic approach to activity, the session at the Center brings together as many of those who have worked on the unit as is possible. In the case of Quetelet--and probably others--several of these people were soon to leave the role situations they had inhabited. This "one last rallying" around a common task builds project cohesion for those who remain and provides a sense of satisfactory closure for those departing. At the social interaction level, such celebratory occasions are quite important.

The Formative Evaluation Process: Functions and Structures

In spite of the insights gained from our study of the formative evaluation of the Quetelet unit, it must be said that overall our material is "thin." The unit was too good, the process too problem-free to be an ideal case for investigation. It has, however, led us to consider the formative evaluation process, its aspects and possible structures, beyond the specific case of Quetelet.

Formative evaluation involves three process-like aspects: (a) technical critique; (b) problem posing, problem solution, and program improvement; and (c) social validation for involved actors. The first two of these

are manifest, the third involves latent functions of the process. All three are complementary requirements of the process and a delicate balance among them is essential.

The first two of these are temporally and logically sequenced. Actors must move from a critical stance vis-a-vis the object of evaluation to a creative, positive stance. Given this sequence, the form of the technical critique and the way in which its results are conveyed affects significantly the form of the problem solution, the creative aspect. The conduct of both of these, in turn, affects the validity of the process as a mechanism for social validation, though this function is not temporally ordered but is simultaneous with the other two.

It would be possible to structure a formative evaluation process in such a way that the technical critique overwhelmed the social validation possibilities and severely constrained the creative aspects of redesign. Similarly, the creative aspect of the process could so be emphasized as to obviate the effects of technical criticism. This would be the case in which something was so aesthetically pleasing to the designer that it was retained in the curriculum in the face of obvious negative evidence regarding its inclusion. Finally, the process could be overcommitted to validation and confirmation and could turn into a mere celebration, ignoring critique and redesign.

Acknowledging these potential dangers in the process, the question arises as to the appropriate structure for a formative evaluation. We do not wish to imply that one, and only one, structure is adequate. However, whatever structure is adopted must take the concerns we have outlined into account. One possibility that occurred to us is that, given its close connections with other creative aspects of design work, formative evaluation should be congruent with these. That is, in the case we have looked at, where a holistic approach is taken to both curriculum design and project interaction, an atomistic approach to formative evaluation would be undesirable and

inappropriate were it indeed possible. This kind of problem is not likely to arise when developers conduct formative evaluation. But where those two functions are fulfilled by different people, it becomes advisable for the evaluators to be sensitive to the designer's "style" in such matters.

In any formative evaluation, structure will vary with the answers to these questions:

1. Who controls the process--developers or evaluators?
2. Which organizations are involved in which functions?
3. In what manner is the data from technical critique fed into the creative redesign process?

Whatever social design is chosen for formative evaluation, it must be one which enhances creativity while making criticism clear. There is a close connection between creativity and validation; and a serious problem, we suspect, is to allow critique to overwhelm the other two.

Another aspect of formative evaluation is the necessary articulation of at least several frames of reference--at minimum, those of the designer, the evaluator, the classroom teacher, and the students in the classroom. Individuals' frames of reference are their means of orienting themselves to the realities they encounter, of constructing the meanings of events. Among this variety of actors, each possesses a frame of reference, some expert and discipline-based, others less reflective. But for each there is a set of values, priorities, and confirming tests, so that variation among them is inevitable. How they are translated one into the other and what their common meeting grounds are to be must therefore be a concern in designing and conducting a formative evaluation.

In the formative evaluation of Quetelet, the question of frame of reference was obviated somewhat by the fact that in the prototype phase, the IS co-directors conducted the process under their own aegis and within their own model of activity. Additionally, Sister Martha provided Quetelet with

a classroom teacher whose view differed little from the IS model. We feel that, as complex and rich a program as is IS, the formative evaluation procedures that have been developed are well suited to the cognitive styles of the dominant actors and the curriculum itself.

We close with a recommendation. There is a strong argument to be made for the inclusion of drastic variations in perspective at the formative rather than summative stage. That is, it might be of great interest, and highly productive for problem detection, to attempt even prototype formative evaluation in a setting or with personnel seemingly quite unsuited to the IS model. Since this program is designed to be disseminated nationally, and it has in fact been introduced to over 150,000 children already, clearly there will be implementation variations among classrooms. Subjecting the program at the prototype evaluation stage to severe "outside" critique might highlight problems which remain hidden in environments such as the one in which the Quetelet program was initially evaluated.

We think that this kind of risk taking can only be undertaken when a store of confidence in the product and its style of design already exists. Such confidence clearly exists regarding IS, and in later formative evaluations, or in iterations of the curriculum for other purposes or settings, such risk taking ought to be considered.

References

Champagne, A. B., & Klopfer, L. E. Formative evaluation in science curriculum development. Journal of Research in Science Teaching, 1974, 11(3), 185-203. (Also LRDC Publication No. 1974/25)

Champagne, A. B., & Klopfer, L. E. Individualized learning in Individualized Science. Unpublished manuscript, University of Pittsburgh, Learning Research and Development Center, 1975.

Cronbach, L. J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.

Leinhardt, G. A strategy for program evaluation. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1975. (Publication No. 1975/16)

Scriven, M. The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), Perspectives in curriculum evaluation. Chicago: Rand McNally, 1967.